EEGLE Data Management

Procedures for Archiving and Retrieving--
Where's my data and How do I get it?

Gregory A. Lang
and
David J. Schwab

NOAA/GLERL
2205 Commonwealth Blvd.
Ann Arbor, MI 48105

Introduction

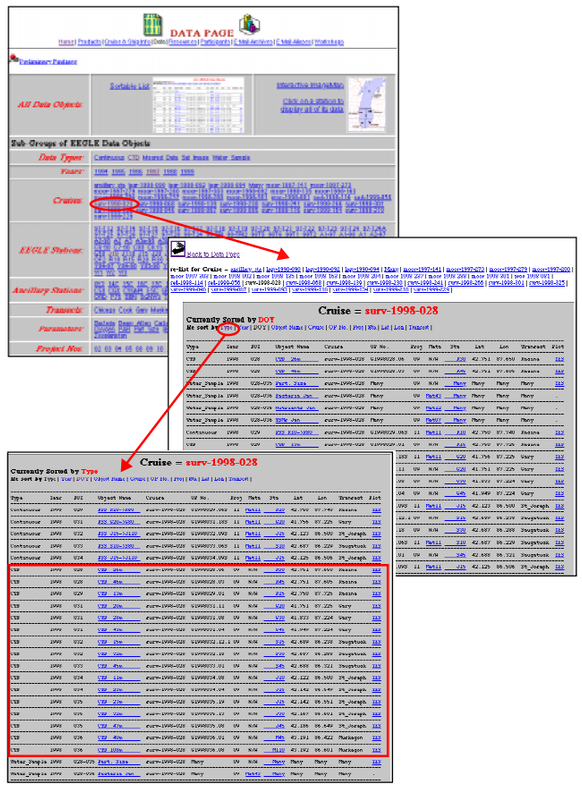
The Episodic Events - Great Lakes Experiment (EEGLE) project will consist of 3 years of field work and 2 years of analysis; with modeling and satellite studies throughout. EEGLE is expected to generate a great deal of measured and modeled data as well as a wide variety of graphical products. One of the goals of the project is to develop a stand-alone database management system to archive and retrieve EEGLE data sets and their descriptive meta files. Web browsers will be used as the standard user interface because of their platform-independent functionality and general availability.

Methods

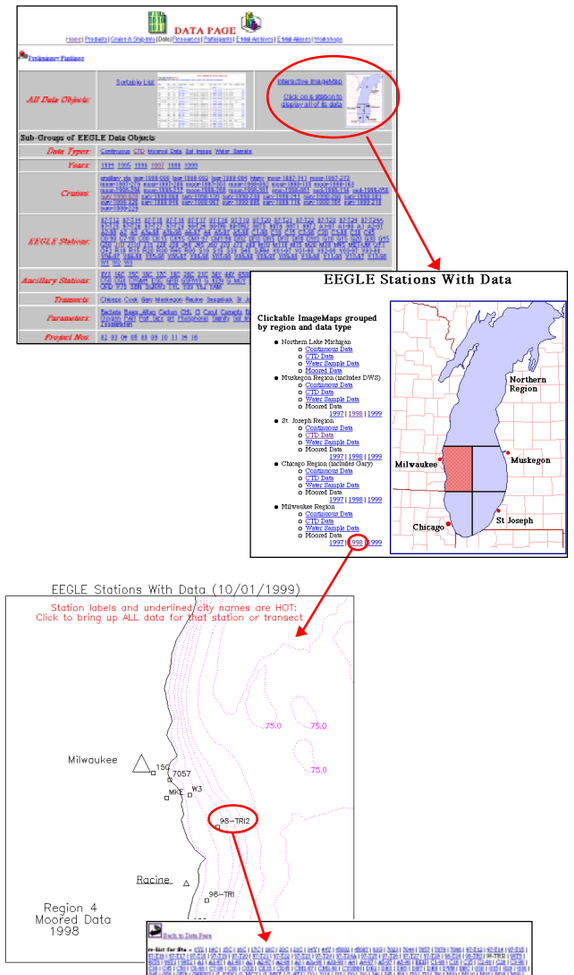
Each data set is entered into the EEGLE database along with its database properties. These properties include: data type (CTD, Moored Data, etc.), year, days of year, object name, cruise, Op No., project no., meta file name, data file name, graphics file name, and measured parameters. From the Op. No., the station name and location are identified.

A database management system, utilizing Unix Bourne shell scripts, IDL programs, and various look-up tables, was developed to access the EEGLE database and generate a series of "HTML" pages suitable for any web browser. These pages are sortable and contain hot links to the data file, the meta file, the cruise plot, and any plots associated with the data.

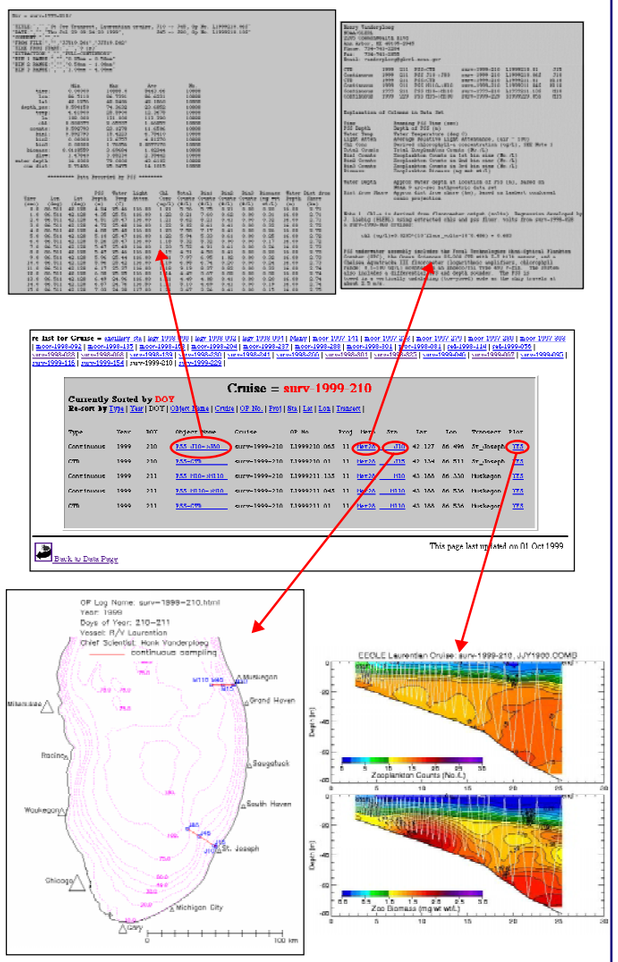
Data can be accessed any number of ways. The user can select the entire EEGLE data set or any sub-group of data. The sub-groups include Data Type, Year, Cruise, Station, Transect, Parameter, and Project No. Each group of data is automatically sorted by year and DOY. The user can choose to re-sort the page to help locate data of interest. For example, to examine all of the CTD data from the surv-1998-028 cruise, the user could first select the sub-group Cruise = surv-1999-028 from the main data page; then re-sort by Type. The CTD files will be aligned together.



In addition, data can be accessed via a series of interactive "ImageMaps." For manageability, the data in the ImageMaps are divided into regions (Northern, Muskegon, St. Joseph, Chicago, and Milwaukee) and data type (Continuous, CTD, Moored, and Water Sample). The Moored Data is further sub-divided by year. Each ImageMap shows the name and location of all stations that contain data. Each station name is "hot" and linked to all of the data sets for that station.



Once a data set is identified, the user can view the data, the meta file describing the data, a plot of the cruise track showing the data collection site, or plots of the data if they exist.



Data Managers' Wish List…

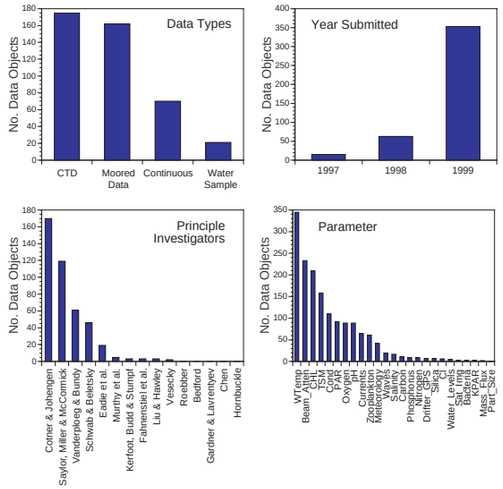
- Submit data in a timely manner.
- Submit Cruise Op Logs in a timely manner.
- Include an Op No. in all data sets and/or meta files.
- Include a descriptive meta file with all data sets.
- Include any plots of the data.
- Try to use uniform format for similar data.
- Space-delimited ASCII is preferred.
- GMT.

Results

To date, 431 unique data sets have been submitted and entered into the EEGLE database. Of these, 175 are classified as CTD, 162 as Moored Data, 70 as Continuous Data, and 21 as Water Sample Data. The number of data sets submitted each year has steadily increased over time; 15 in 1997, 63 in 1998, and 353 in 1999. Water temperature is the most commonly measured parameter, followed by beam attenuation, chlorophyll, TSM, currents, zooplankton, meteorology, and waves. The amount of traffic or "hits" on the EEGLE web site has steadily grown from about 1100 per month in January 1998 to just over 5000 in September 1999. The EEGLE Data area receives one-third of all EEGLE traffic.

The database program developed for the EEGLE project has performed well. It was designed to create all of the pre-sorted HTML pages and ImageMaps each time a data set is entered into the database, thus easing the strain on EEGLE's web server and greatly reducing access time for data pages. The data and plots within the EEGLE Data area now total 450 MB. It is anticipated that this number will grow substantially, with the addition of a third full field year and the inclusion of the modeling results. During the formation of the EEGLE data policy, it was anticipated that all EEGLE data would fit onto a single CD-ROM. This appears unlikely. One possible solution would be to store all biological data on one CD-ROM, all modeling results on another, and possibly all moored instrument data on a third.

EEGLE Data Objects Submitted



No. Hits on EEGLE Web Site